Estimating the Impacts of Program Benefits:

Using Instrumental Variables with

Underreported and Imputed Data*

Melvin Stephens Jr.       Takashi Unayama

University of Michigan   Policy Research Institute
and NBER                  Ministry of Finance

This Version: April 4, 2014

*Stephens: Department of Economics, University of Michigan, 611 Tappan St., 341 Lorch Hall, Ann Arbor, MI 48109-1220, e-mail: mstep@umich.edu and National Bureau of Economic Research, Cambridge, MA. Unayama: Policy Research Institute, Ministry of Finance, 3-1-1 Kasumigaseki, Chiyoda-ku, Tokyo, Japan, e-mail: takashi.unayama@mof.go.jp. We thanks seminar participants at Michigan and Williams for helpful suggestions. We also thank the Statistical Bureau of the Japanese Government for allowing access to the Family Income and Expenditure Survey data. The views expressed in this paper do not necessarily reflect the views or policy of the MOF.
Estimating the Impacts of Program Benefits:
Using Instrumental Variables with Underreported and Imputed Data

Abstract

Survey non-response has risen in recent years which has generated an increase in the share of imputed values found on commonly used datasets. While the importance of this trend for earnings is well-known, the growth in imputed values for government transfers has received far less attention as has the underreporting of transfer benefits. We demonstrate analytically that underreporting and imputation of government transfers can lead to estimates of program impacts that are substantially overstated when applying instrumental variables methods to correct for endogeneity and/or measurement error in these measures. Our theoretical predictions are borne out in both of our empirical examples.
1 Introduction

Vast economic literatures estimate the impacts of government benefits on an array of household-level and individual-level outcomes (e.g., see the surveys of Krueger and Meyer (2002) and Currie (2004)). Since participation in these programs is quite often a choice, a typical empirical approach for estimating the causal effect of benefit receipt is to treat the amount of benefits received as an endogenous regressor and then implement an estimation method such as instrumental variables (IV) that exploits an exogenous source of variation in benefits. Although benefit amounts reported on household surveys are typically thought to be measured with error, it is well-known that IV estimates are consistent if the measurement error in the endogenous regressor is classical. However, it is quite common for benefit amounts on surveys to be underreported or for the data provider to replace missing information with imputed values. In this paper, we demonstrate both analytically and through the use of two empirical examples that IV estimation with underreported or imputed endogenous variables will tend to overstate, and in some instances quite substantially, the causal effect of program benefits.

Household values of program benefits received found on a number of major surveys such as the Current Population Survey (CPS), the Decennial Census, the Panel Study of Income Dynamics (PSID), and the Survey of Income and Program Participation (SIPP) suffer from measurement error that is non-classical. First, when households acknowledge receiving a benefit but do not recall the amount, these benefits routinely are imputed. While it is well-known that a substantial and increasing number of earnings reports are imputed (Lillard, Smith, and Welch (1986); Hirsch and Schumacher (2004); Bollinger and Hirsch (2006)), the extent of benefit imputation has received far less attention. Figure 1 displays the share of imputed income measures among respondents with non-zero amounts for each measure in the March CPS from 1988-2013. The Figure shows not only that benefit imputations have increased just as dramatically as earnings over this period. The imputation rates of positive earnings reports are only slightly larger than those of positive Supple-

---

1Meyer, Mok, and Sullivan (2009) present a similar set of results although in terms of dollars imputed rather than individuals. Prior to 1988, unemployment insurance and worker’s compensation benefits are combined with other government benefits and cannot be separately examined.

2Determining which earnings values are imputed in the March CPS becomes less transparent beginning in 1988. A discussion of the issues involved and the necessary Stata and SAS code can be found at http://www.psc.isr.umich.edu/dis/data/kb/answer/1349.
mental Security income reports while Social Security imputation rates exceed those of earnings.\(^3\)

A second issue is that benefit levels are underreported in surveys. For example, the Consumer Expenditure Survey (CE) did not construct imputations prior to 2004 although roughly one in six of the consumer units were flagged as “incomplete income reporters.” However, only one valid non-zero report for a single major source of income was needed to qualify as a complete income reporter which means this designation is likely a misnomer for a number families (Paulin and Ferraro 1994). Meyer, Mok, and Sullivan (2009) find across a number of benefit categories that total benefits received, computed by aggregating and appropriately weighting responses in a given survey, yields benefit levels that fall short of administrative records detailing the total amount of benefits paid. This total dollars underreporting occurs even after accounting for imputations. Thus, benefit imputations may fall short of actual benefits received for those individuals that acknowledge receipt but forget amounts received or some households may fail to report the receipt of benefits (or both).

In this paper, we show that underreporting and imputation can lead instrumental variables estimates to dramatically overstate the impacts of transfer programs. Hirsch and Schumacher (2004) and Bollinger and Hirsch (2006) demonstrate that regression coefficients will be attenuated when using imputed outcomes if the imputation procedure does not explicitly condition on the regressor of interest. For example, they note that the CPS imputation method does not account for union status when assigning earnings to those that fail to report an earnings amount. They then show that union wage gaps are substantially understated when using the full CPS sample as compared to the sub-sample of observations with non-imputed earnings information.

When estimating the impact of transfer programs, the amount of benefits received is typically treated as an endogenous regressor. As part of the IV methods used to account for this endogeneity, the benefits received will be the outcome in the first stage equation. If the imputation method used to replace to missing benefit amounts does not condition on the instrumental variable, the estimated first stage coefficient on the instrument converges to a value smaller than the true parameter as in

\(^3\)These imputation rates not only account for the item non-response but also for the fact that each year roughly 10% of households do not answer (or only provided limited data for) the March CPS supplement. Whole supplement non-response information is contained in the CPS variable FL-665. While the whole supplement non-response flag does not appear on the public-use CPS data until 1991, it does appear on the Unicon CPS files beginning in 1988. We thanks Jay Stewart of the Bureau of Labor Statistics for directing us to these pre-1991 data.
the case of union wage gaps. For example, if the instrument relies on changes in benefit program rules that vary across states and over time but the imputation procedure does not condition on state of residence, then the imputed values will not be correlated with the instrumental variable.

Since the IV estimate can be computed as the ratio of the reduced form to the first stage coefficients on the instrument, the shrinking of the first stage estimate due to this inconsistency will yield an IV estimate that exceeds its true parameter value. We show analytically that when the instrument is uncorrelated with the imputed values, the IV estimate will converge to a value exceeding the true IV parameter by a factor of $1/p$ where $p$ is the fraction of households correctly reporting benefit values. Given that only two-thirds of CPS households correctly report benefits in recent years, IV estimates that use imputed benefits will be biased upwards by 50 percent.

The underreporting of benefits yields a similar impact on the IV estimate. If underreporting takes the form of not reporting benefits at all then the resulting IV estimator is the same as in the situation when the imputations are uncorrelated with the instrumental variable. The extent of this inconsistency falls as the degree of underreporting diminishes as we demonstrate below.

To illustrate the impact of these survey non-response issues, we present two empirical examples. The first example makes use of the well-known U.S. Social Security “notch.” Social Security benefits were adjusted to automatically account for inflation beginning in 1975 rather than having Congress make such adjustments on an ad hoc basis. However, the first attempt at regular cost of living adjustments actually led to a “double indexation” of benefits which led to a rapid growth in benefits for new benefit recipients in the years that followed. Congress ultimately decided not to reduce benefits for those that had begun receiving benefits already but rather adjusted the formula so that benefits were reduced gradually beginning with the 1917 birth cohort and the double indexation was entirely eliminated with the 1921 birth cohort.

We follow Englehardt, Gruber, and Perry (2005) by using the notch to examine the impact of Social Security income on the ability of elderly households to live independently. However, since Social Security benefits are not imputed with respect to exact age in the CPS, but rather by using broad age categories, the impact of the notch-based instrument on Social Security benefits is four times larger for families with non-imputed benefits than those with imputed benefits. We find that

---

4Krueger and Pischke (1992) discuss the details of the legislative changes surrounding the notch.
the IV estimate of the impact of Social Security benefits on living arrangements is biased upwards 20 to 30 percent.

Our second example examines the consumption response of Japanese households to child benefit payments which arrive once every four months. These benefits are severely underreported in the data as only one quarter of eligible households report receiving these payments, perhaps due to the infrequency with which the benefits are received. A common approach would be to instrument for household benefits with the known benefit payment pattern which is based on calendar month and the age and number of children in the household. Since monthly benefits are underreported, specifically with regards to benefit payments, we demonstrate that the IV estimator is dramatically inconsistent in the upwards direction and produce estimates that are overstated by a factor of four. Regressing the change in consumption on the programmatic change in monthly child benefits (i.e., the reduced form) or dropping households that do not report benefits (under the assumption that benefits are missing at random) yields substantially smaller benefits as predicted.

While IV methods can be used when regressors are measured with error, the errors must follow the classical measurement error model to yield consistent estimates. Underreporting is not a form of classical measurement error while imputations are not either unless the imputation procedure accounts for all possible determinants of the missing variable. The measurement error induced by underreporting and imputations is akin to “mean-reverting” measurement error. Berger, Black, and Scott (2000) analyze the inconsistency of the IV estimator when using one noisy measure to instrument for another noisy measure when both are contaminated with mean reverting measurement error. In our analysis, the inconsistency of the IV estimator arises when the endogenous regressor is either underreported or imputed even when the instrument is perfectly measured. Our analysis is appropriate for many settings where program benefit rules vary by well-measured observable characteristics such as age and state of residence such that there is typically little error in the instrumental variable. Our analysis can easily be extended to situations in which the outcome of interest is also underreported and/or imputed.
2  Econometric Framework

We focus on the population regression model for a continuous outcome $y$

$$y = \beta_0 + \beta_1 x + u$$  \hspace{1cm} (1)$$

where $x$ is an endogenous, continuous regressor such that $\text{Cov}(x, u) \neq 0$. It is straightforward to extend our analysis to include additional exogenous covariates as we show below.

Suppose that $z$ is a valid, continuous instrumental variable for $x$ such that $\text{Cov}(x, z) \neq 0$ and $\text{Cov}(z, u) = 0$ and the population regression model for $x$ is

$$x = \pi_0 + \pi_1 z + \epsilon$$  \hspace{1cm} (2)$$

while the corresponding reduced form equation is

$$y = \delta_0 + \delta_1 z + \epsilon$$  \hspace{1cm} (3)$$

When the data are free of measurement error, the OLS estimators for the coefficients on $z$ in equations (2) and (3), $\hat{\pi}_1$ and $\hat{\delta}_1$, respectively, are consistent since $z$ is assumed to be exogenous.

It is well-known that the instrumental variables (IV) estimator for $\beta_1$ can be written as the ratio of the OLS estimators of the slopes in the reduced form to the first stage equations

$$\hat{\beta}_{1IV} = \frac{\sum_i (z_i - \bar{z}) (y_i - \bar{y})}{\sum_i (z_i - \bar{z}) (x_i - \bar{x})}$$

$$= \frac{\left( \sum_i (z_i - \bar{z}) (y_i - \bar{y}) \right) / \left( \sum_i (z_i - \bar{z})^2 \right)}{\left( \sum_i (z_i - \bar{z}) (x_i - \bar{x}) \right) / \left( \sum_i (z_i - \bar{z})^2 \right)}$$  \hspace{1cm} (4)$$

$$= \frac{\hat{\delta}_1}{\hat{\pi}_1}$$

Typically the amount of government benefits received is the endogenous regressor when examining the impact of a program on an outcome of interest. However, the instrumental variables
used in the analysis typically are based on demographic characteristics that are far less likely to be misreported. For example, Medicaid eligibility is driven primarily by the age of a child while the maximum earned income tax credit (EITC) depends upon the number of children in the household. As such, it is quite common to arrive in a situation where the endogenous regressor is underreported or imputed while the instrumental variable is not.

Suppose we can group observations based upon whether the endogenous regressor, \( x \), is an actual report, \( A \), or an underreport/imputed values, \( U \). As is well-known, we can write the OLS estimator as a weighted average of the OLS estimators for each sub-group. Applying this result to the first stage slope coefficient yields

$$
\hat{\pi}_1 = \frac{\sum_i (z_i - \bar{z})(x_i - \bar{x})}{\sum_i (z_i - \bar{z})^2} = \frac{\sum_{i \in A} (z_i - \bar{z})(x_i - \bar{x}) + \sum_{i \in U} (z_i - \bar{z})(x_i - \bar{x})}{\sum_i (z_i - \bar{z})^2} = \frac{\sum_{i \in A} (z_i - \bar{z})^2 \cdot \sum_{i \in A} (z_i - \bar{z})(x_i - \bar{x}) + \sum_{i \in U} (z_i - \bar{z})^2 \cdot \sum_{i \in U} (z_i - \bar{z})(x_i - \bar{x})}{\sum_i (z_i - \bar{z})^2} \cdot \frac{\hat{\pi}_{1\text{, }A}}{SS_{z\text{, }A}} + \frac{SS_{z\text{, }U}}{SS_{z}} \cdot \hat{\pi}_{1\text{, }U}
$$

Thus, the OLS estimator for the first stage slope coefficient, \( \hat{\pi}_1 \), is a weighted average of the corresponding estimators when the model is estimated separately for each group, \( \hat{\pi}_{1\text{, }A} \) and \( \hat{\pi}_{1\text{, }U} \), where the weights are the share of the variation in the instrument, \( SS_z \), belonging to each group.

Suppose that whether \( x \) is an actual or an under/imputed report is randomly assigned where \( p \) is the probability of providing an actual report. The first stage slope estimator for the sample of actual reporters, \( \hat{\pi}_{1\text{, }A} \) will be a consistent estimator for \( \pi_1 \). In addition, the weights in the final line of (5), \( \frac{SS_{z\text{, }A}}{SS_z} \) and \( \frac{SS_{z\text{, }U}}{SS_z} \), are consistent estimators of \( p \) and \( 1 - p \), respectively. However, the corresponding estimator for the under/imputed reporters, \( \hat{\pi}_{1\text{, }U} \), depends upon the corresponding underreporting or imputation process.

When the endogenous regressor is imputed, the impact on \( \hat{\pi}_1 \) depends upon the imputation
process. For example, the “hot deck” imputation procedure used by the U.S. Census Bureau to allocate values when items are missing due to non-response selects a replacement amount from a “donor” who has the same values for a small set of observed characteristics. While this procedure retains the covariance between the allocated variable and the characteristics used as part of the matching process, Hirsch and Schumacher (2004) and Bollinger and Hirsch (2006) note that the covariance between the allocated variable and other characteristics is not preserved. If the procedure to impute missing values for the endogenous regressor \( x \) does not depend upon the instrumental variable, then there will fail to be a correlation between \( x \) and \( z \) among the imputed observations. Thus, \( \hat{\pi}_{1,U} \approx 0 \) and the probability limit of the first stage slope estimator \( \hat{\pi}_1 \) will equal \( p\pi_1 + (1 - p)\pi_1 = p\pi_1 \).

In the case of underreporting, suppose that the underreported value of \( x \) is a constant fraction \( \theta \) of actual \( x \) among all underreporters. It is straightforward to show that the probability limit of the first stage slope coefficient for underreporters is \( \theta \cdot \pi_1 \). Thus, with underreporting, the probability limit of the first stage estimator is \( p\pi_1 + (1 - p) \theta \pi_1 = \pi_1 (p + [1 - p] \theta) < \pi_1 \).

In the extreme case where individuals simply forget to report \( x \) (i.e., \( \theta = 0 \)), the probability limit of \( \hat{\pi}_1 \) falls to \( p\pi_1 \). Failure to report benefits might arise when benefits are small in value or are received infrequently. Our analysis directly translates to some instances where benefits are not reported, e.g., if program benefits are universal but yet respondents fail to report receipt of these payments. In other instances, we may simply lack knowledge as to whether or not recipients in fact took up these benefits. For example, Blank and Card (1991) find take-up rates of roughly two-thirds among those eligible for Unemployment Insurance (UI) benefits. However, Meyer, Mok, and Sullivan (2009) find that roughly 75% of UI benefit payments are reported in a number prominent national surveys even when including imputed benefit payments. If under reporters are treated as failing to take up benefits, then the issues raised here may be quite important for interpreting the findings.

The impact of underreported and imputed values of \( x \) on the IV estimator can be seen by
substituting in (5) and an analogous expression for the reduced form estimator into (4)

\[
\hat{\beta}^{IV}_1 = \frac{\hat{\delta}_1}{\hat{\pi}_1} = \frac{SS_{z,A} \cdot \hat{\delta}_{1,A} + SS_{z,U} \cdot \hat{\delta}_{1,U}}{SS_{z} \cdot \hat{\pi}_{1,A} + SS_{z} \cdot \hat{\pi}_{1,U}}
\]  

(6)

As we discussed above, the denominator will converge in probability to values smaller than \(\pi_1\) when the endogenous regressor is either underreported or imputed. As long as the dependent variable in the main equation, \(y\), is not imputed or underreported, the reduced for slope coefficients, \(\hat{\delta}_{1,A}\) and \(\hat{\delta}_{1,U}\), for the actual and under/imputed reporters, respectively, will yield consistent estimates of \(\delta_1\). Thus, the probability limit of the IV estimator will exceed \(\beta_1\) due to the underreporting or imputation of the endogenous regressor. In cases where either underreporters all state that they received no benefit or where the imputations are uncorrelated with the instrumental variable, the probability limit of \(\hat{\beta}^{IV}_1\) becomes \((1/p) \beta_1\).

Extending the analysis to include exogenous regressors is straightforward through an application of the Frisch-Waugh-Lovell Theorem. As is well-known, the OLS estimate for the coefficient on \(x\) when regressing \(y\) on \(x\) plus a vector of covariates \(w\) is numerically equivalent to performing the following steps: a) regressing \(y\) on \(w\) and saving the residuals, \(y^*\), b) regressing \(x\) on \(w\) and saving the residuals, \(x^*\), and c) regressing \(y^*\) on \(x^*\). Similarly, the 2SLS estimate for the coefficient on \(x\) when regressing \(y\) on \(x\) plus a vector of covariates \(w\) using \(z\) as an instrument for \(x\) is numerically equivalent to first separately regressing \(y\), \(x\), and \(z\) on \(w\) and saving the residuals, \(y^*\), \(x^*\), and \(z^*\), respectively, and then regressing \(y^*\) on \(x^*\) using \(z^*\) as an instrument. Thus, we can continue to apply (6) except that the reduced form and first stage coefficients are based on the regressions using \(y^*\), \(x^*\), and \(z^*\) and the weights depend upon the shares of the variation in \(z^*\) between the actual and underreported/imputed sub-samples.
3 Empirical Examples

3.1 The Impact of the Social Security Notch Using Imputed Social Security Benefits

The U.S. Social Security “notch” generated a sizable change in Social Security benefits for the affected birth cohorts.\(^5\) Social Security benefits were adjusted to automatically account for inflation beginning in 1975 rather than having Congress make such adjustments on an ad hoc basis. However, the first attempt at regular cost of living adjustments actually led to a “double indexation” of benefits. The double indexation resulted from using both a non-indexed measure of average monthly earnings that increased with inflation along with having a replacement rate that increased with inflation.\(^6\) Congress ultimately decided not to reduce benefits for those that had begun receiving benefits already but rather adjusted the formula so that benefits were reduced beginning with the 1917 birth cohort and the double indexation was entirely eliminated with the 1921 birth cohort.\(^7\)

Englehardt, Gruber, and Perry (2005) (EGP) use the notch to investigate the impact of Social Security income on the probability that elderly-headed families are living independently. OLS regressions of this relationship are almost certainly inconsistent. Since Social Security benefits are a function of lifetime earnings, differences in living arrangements are more likely to reflect permanent income differences across households than the causal influence of Social Security income.

EGP use data from the 1980-1999 March CPS supplements to analyze this issue. They limit their analysis to families where the Social Security recipient is age 65 and older in the case where this recipient is male or a never married female and ages 62 and up in the case of a widowed or divorced female.\(^8\) They also limit the analysis to the 1900-1993 birth cohorts which are roughly symmetric around the 1916 cohort for whom the Social Security reached a peak before the laws generating the notch became effective.

EGP create their instrument for the impact of the notch on Social Security benefits by entering

\(^5\)Krueger and Pischke (1992) discuss the details of the legislative changes surrounding the notch.

\(^6\)The current formula first indexes average monthly earnings before computing before benefit amount but does not index the replacement rate.

\(^7\)The Social Security Administration’s explanation of the notch is found at http://www.ssa.gov/history/pdf/notch.pdf.

\(^8\)They do not use the CPS definition of families within households but rather treat the primary CPS family as one family and all other adults in the household as individual families. See their paper for more details of this procedure.
the same earnings profile into a Social Security benefits for different year of birth cohorts. They fix the wage profile to reflect the median male earner in the 1916 birth cohort. They use this profile not only to compute the Social Security benefit of someone born in 1916, but also for every birth cohort from 1900-1933 by using the Consumer Price Index (CPI) to inflate wages across time. By using the wage profile for a single birth cohort, the idea is that the instrumental variable will only reflect changes in the programmatic rules across birth cohorts. When the Social Security recipient is a widowed or divorced female, the benefit is assumed to be based on her deceased/ex-husband’s earnings record. Since they find through other data sources that these husbands were on average three years older, they base the instrument on the birth cohort three years older than the woman’s birth cohort.

The solid line in Figure 2 shows the profile by year of birth cohort. Benefits rise rapidly for birth cohorts up to and including the 1916 birth cohort. Due to the correction for double indexation discussed above, the benefits fall rapidly for the 1917-1921 birth cohorts before remaining relatively flat for the remaining birth cohorts displayed in the figure.

As shown previously in Figure 1, roughly 20% of Social Security benefit recipients in the CPS have imputed benefits while this figure steadily rises to nearly 30% of recipients by 1999. Imputations arise from two sources: item non-response and whole supplement non-response. Item non-response occurs when the respondent indicates that they have received, e.g., Social Security benefits, but do not recall or disclose the amount of benefits received. Whole supplement non-response occurs when households finish the basic CPS interview but then refuse to participate in the March Supplement. Whole supplement non-response has remained constant at roughly ten percent such that the increase in non-reporting in recent years is driving by item non-response.

As we discussed above, the CPS uses the “hot deck” imputation method to allocate missing values by taking a value from a donor observation with the exact same set of characteristics. For

---

9 EGP assume that benefits are claimed beginning at age 65. They provide more details regarding the construction of the instrument in their paper.

10 We thank Gary Englehardt for sharing the values of the instruments by birth cohorts. The data actually contain four series for the instrument computed separately for high school dropouts, high school graduates, those attending college but not completing, and those with college degrees. The Figure displays a weighted average of these four series based on the distribution of Social Security recipients in our sample.

11 Prior to 1988, information on whole supplement imputation was contained in the data allocation flag for each income measure.
whole supplement non-response, the entire set of responses to the supplement are taken from a donor. To get an exact match, continuous variables are collapsed into categorical values while some values are a single categorical characteristic are combined.

In the case of item non-response for Social Security benefits, the characteristics used in the match are age (7 categories), sex (2), marital status (4), race/ethnicity (3), education (3), work status (2), and pension type (6) resulting in 6,048 possible combinations.\textsuperscript{12} If a donor with an exact categorical match is not found, the race/ethnicity and work status categories are dropped from the procedure and a donor is pursued from the resulting 1,008 combinations.\textsuperscript{13}

Most relevant for our analysis is that the method used allocate missing Social Security benefits due to item non-response is not based on exact year of birth/age. Instead, the seven age categories used in this procedure are: less than 35, 35 to 54, 55 to 61, 62 to 64, 65 to 69, 70 to 74, and 75 and over. Since our analysis is restricted to those ages 65 and older (62 and older in the case of widowed and divorced women), there are only three (four) relevant age categories used in the imputation procedure. As such, the imputed benefits due to item non-response will fail to capture the sharp spike in Social Security income by birth cohort that is exhibited by the notch instrument.

When the entire supplement is missing, the matching procedure depends first upon marital status and labor force status. Within the five possible marital status/employment groupings, age is again used a match category. However, ages 65 and up are always grouped together, and in some cases ages 55 and up pooled together if a donor is not initially found. Thus, whole supplement non-response will also fail to capture the year of birth based movements in Social Security benefits due to the notch.

The long and short dashed lines in Figure 2 show the average Social Security income by birth cohort for non-imputed and imputed values, respectively. Whereas the actual reports of Social Security income in the CPS exhibit strong evidence of the notch, there is no evidence of a notch among the imputed values. There certainly is evidence of an increase in benefits for the pre-notch cohorts which is consistent with the fact that the imputations are based on somewhat close ages.

\textsuperscript{12}We thank Ed Welniak for providing us with the internal Census Bureau documents detailing the characteristics used in the hot deck procedure.

\textsuperscript{13}According to the internal Census documents, in the “rare occurrence” that a donor is not found from this second pass at generating an allocated value, an amount “will be imputed from a matrix.”
However, the imputation procedure does not capture the rapid decline in benefits following the implementation of the notch.\footnote{Although not shown here, when the imputed values are separated between item non-response and whole imputation non-response, neither series shows any evidence of a decline following 1916.}

EGP estimate a model of the form

\[ P_{i,t} = \theta SSIncome_{i,t} + \beta X_{i,t} + \gamma_i + \alpha_t + \phi_i + u_{i,t} \]  \hspace{1cm} (7)

where \( P_{i,t} \) is an indicator for having a shared living arrangement; \( SSIncome_{i,t} \) is family Social Security income measured in thousands of dollars; \( X_{i,t} \) includes four categories each for the education of the head and the spouse (if present), age of the spouse (if present), marital status (married, widowed, and divorced), white, and female; \( \gamma_i \) is a full set of indicators for the age (age+3 for widowed and divorced women) from ages 65 to 90; \( \alpha_t \) is a set of survey year indicators, and \( \phi_i \) is a set of indicators for the nine Census divisions. The legislative Social Security benefits based on the median earner from the 1916 birth cohort shown in Figure 2 is used as an instrument for \( SSIncome_{i,t} \).

We should note some important differences between our analysis and EGP’s. First, we currently only have the Social Security instrument for the 1900-1928 birth cohorts rather than through the 1933 birth cohort. Second, we are using the instrument based on the four education groups as opposed to a single value for each birth cohort. Third, whereas EGP create age-by-year of birth cells for each survey year, we use individual level data although this last difference should create any substantive differences.

Table 1 presents our results.\footnote{All of our estimates are weighted by the individual sampling weight for the Social Security recipient. The standard errors are clustered at the year of birth level.} Based on our discussion above, we apply the Frisch-Waugh-Lovell theorem by first regressing the shared living indicator, Social Security income, and the instrument on the remaining exogenous covariates in the model and saving the residuals. We then use these residuals to estimate the first stage and reduced form models across both the full sample and the two sub-samples to be constant with the decomposition shown in equation (4). The OLS and 2SLS results do not use the residuals but rather are estimated separately for the full sample as well as for each sub-sample.
Panel A in Table 1 presents our findings for the full sample of 248,075 families in column (1) whereas are findings for those without and with imputed Social Security income are shown in columns (2) and (3), respectively. Applying OLS to equation (7) finds that the probability of living in a shared arrangement falls with increases in Social Security. The impact is over 250% larger in the non-imputed sample than in the imputed sample which is consistent with attenuation bias due to imputation reducing the estimate for those with allocated benefits.

The first stage estimates vary across the columns as predicted by our analytical results. The estimated effect of the instrument on Social Security income is nearly 20% larger in the non-imputed sample than in the full sample, consistent with the share of Social Security benefits that are imputed in the CPS during this period. In addition, the estimated first stage relationship is more than three times larger for the non-imputed sample than the imputed sample which is consistent with the inability of the imputation proceed to match the sharp changes in benefits due to the notch.

Also as anticipated, there is little variability in the reduced form estimates across the three columns. The shared living arrangements measure is based strictly on the household roster which is far less likely to be affected by measurement error than Social Security income. There is no reason to expect that the relationship between this variable and the instrument would systematically vary based on whether or not Social Security income is imputed.

The final row of Panel A shows the 2SLS estimates of the impact of Social Security income for the full sample as well as separate estimates for the non-imputed and imputed samples. We see that the 2SLS estimate of -0.36 is nearly 25% larger when using the full sample as opposed to the estimate of -0.29 using the non-imputed sample only. Assuming that Social Security benefits are missing at random, the results from the full sample substantially overstate the efficacy of Social Security benefits in reducing shared living arrangements.

EGP also compute elasticities by combining the point estimate along with the sample fraction living in shared arrangements and average Social Security income. Given that 25.0% of the full sample is in shared living arrangements and average Social Security income is $5,774, and remembering that the income is measured in thousands of dollars in the regression, yields an elasticity of -0.83.\textsuperscript{16} Doing the same calculation for the non-imputed sample lowers the elasticity to -0.66 which

\textsuperscript{16}EGP find an elasticity of -0.41. Our 2SLS estimate of -0.036 whereas their 2SLS estimate is -0.0205. We are
means that the inclusion of the imputed values raises the estimated elasticity by over 25%.

In the final column of Table 1, we present results that account for the fact that the non-imputed observations may constitute a selected sample. Following Bollinger and Hirsch (2006), we use inverse propensity score weighting (IPW) to correct for sample selection based on observable characteristics. We estimate a probit for using an indicator for reporting an actual Social Security value as the dependent variable using the same regressors as we include in (7). We then re-estimate all of the equations for the non-imputed sample using the IPW based on the probit estimates. As shown in column (4),

Panel B presents results for the sub-sample of widowed families for whom EGP find an elasticity of -1.30. We again find that the first stage estimate for the non-imputed sample is nearly 20% larger than the pooled sample and is now nearly four times as large as the sample with imputed benefits. The reduced form is somewhat larger for the imputed households relative to the non-imputed households although the estimates for the entire sample of widows and those with non-imputed benefits are quite similar.

The 2SLS estimate of -0.146 for the full sample of widows shown in the final row of Panel B is more than 30% larger than the corresponding estimate for the sample of families without imputed Social Security benefits. The corresponding elasticities for the entire sample of widows and those with non-imputed benefits are -2.24 and -1.70, respectively. Again, including families with imputed benefits leads to a greatly overstates the impact of Social Security benefits on shared living arrangements.

3.2 Excess Sensitivity and the Underreporting of Japan’s Child Benefit

Public transfers to families on the basis of the age and number of their children is prevalent in a number of developed countries (OECD 2011). Japan introduced its child benefit system in 1972 although only households with three or more children initially received benefits from this program.
Families with two or children became eligible in 1986 while eligibility was extended to families with one child in 1992. Benefits were initially available until the child was fifteen although the eligibility age was reduced to three when the first child became eligible in 1986. In subsequent years, the age limit has been raised repeatedly, rising to six in 2000, to nine in 2004, to twelve in 2006, and to fifteen in 2009. Child benefits were means tested until 2009 after which only the age and number of children are the criteria for the receiving these payments.

While benefit eligibility changed due to parity and child age over this period, during the 1970s and early 1980s, the benefit level per child remained relatively stable in real terms. Beginning in 1992, benefits were set at five thousand yen per month for first and second child and at ten thousand yen for each additional child. In 2006, the monthly benefit amount was set at ten thousand yen regardless of parity but only until age three. Benefits were significantly increased between 2009 and 2012, but then subsequently decreased. Stephens and Unayama (2014) find that child benefits rise as a share of family income during this period, peaking at over 3.5 percent in the late 1990s and remaining above 3 percent in the years that followed.

While child benefits amounts are given in terms of months above, benefits are only received three times a year, in equal amounts, during February, June, and October. The Life-Cycle/Permanent Income Hypothesis (LCPIH) predicts that households will smooth consumption in response to predictable changes in income including the receipt of regular income. However, a number of papers find that consumption is sensitive to the timing of income receipt including various types of government transfer payments (Stephens 2003; Shapiro 2005, Mastrobouni and Weinberg 2009; Stephens and Unayama 2011) and paychecks (Stephens 2006).

Using monthly data from the Japanese Family Income and Expenditure Survey (JFIES) from 1992-2008, we test whether monthly household consumption responds to the receipt of child benefit payments. We limit our sample to the period where families with one child are eligible to receive child benefits but before the means test was eliminated. Families are surveyed in the JFIES for six consecutive months. Each day during the survey period, households are instructed to enter all expenditures and income into diary. Our sample contains monthly summaries of expenditure.

19 The effective eligible age was 5 in 1972 and 10 in 1973 as a transition.
and income in a detailed set of categories.\textsuperscript{20} We construct a measure of monthly non-durable expenditure to test the LCPIIH as this category is commonly used in the literature testing excess sensitivity (e.g., see Parker (1999) and Souleles (1999)).

Child benefits are included in the data as part of a variable titled “other social security.” This variable contains benefits received from any social welfare program except for public pension payments. In the months that benefits are distributed, we find that 24\% of households report positive benefits for this variable. Among these positive reporters, 70\% report values which exactly match what we expect they would receive based on programmatic rules. Two-thirds of the remaining positive reporting households give values that exceed the eligible value, consistent with additional benefits being received and summed into this variable. In the months in which benefits are not distributed, only 4\% of households report positive benefit receipt which most likely is attributable to other social welfare programs.

To examine whether non-durable consumption is excessively sensitive to child benefits, we estimate the equation

\begin{equation}
\Delta C_{i,t} = \alpha_0 + \alpha \Delta HHI\text{income}_{i,t} + \gamma X_{i,t} + u_{i,t} \tag{8}
\end{equation}

where $\Delta C_{i,t}$ is the change in consumption from month $t - 1$ to month $t$, $\Delta HHI\text{income}_{i,t}$ is the change in household income between adjacent months, and $X_{i,t}$ are additional controls for monthly consumption growth including calendar year and month indicators, survey month indicators, the change in the number of household members, and the age of the household head and its square.

Of course, simply estimating (8) using OLS does not yield a valid test of the LCPIIH since any unanticipated changes in income might also influence consumption decisions. Since consumption should not be correlated with predictable income changes, we can use the anticipated change in income due to the timing of child benefit payments as an instrumental variable for income. We can construct the child benefit instrument based on the programmatic rules which are a function of the age and number of children as well as the means test, i.e., whether or not household income exceeds the annual income threshold to receive child benefits.\textsuperscript{21}

\textsuperscript{20}More detail regarding the JFIES is given in Stephens and Unayama (2011).

\textsuperscript{21}Although households only appear in the JFIES for six months, upon entry into the sample households are asked to report total household income for the twelve months prior to the survey period. We use this income measure to determine whether or not households are above or below the means test threshold.
Since we observe that roughly three-fourths of households do not report receipt of the child benefit payments even though they are eligible for these transfers, we know that the IV estimator will suffer from the inconsistency discussed in the previous section. One potential concern with this interpretation might be that non-reporters may, in fact, be ineligible for child benefits. If we mis-measured benefit eligibility in this way, we would expect higher rates of zero benefit reporting among higher income benefit eligible households since they are more likely to be misclassified. However, as we show below, when we split the sample between high and low income households to examine whether the response can be attributed to liquidity constraints (e.g., Zeldes (1989)), we find nearly identical first stage estimates of the impact of child benefits on income for both above and below median income households.

Our measure of household income, $HH_{income_{i,t}}$, includes all sources of household income including regular earnings, interest payments, government transfers, etc. However, we exclude bonus income from our measure of household income. The two months in which bonuses are typically received in Japan are December and June, with June also being a month in which child benefits are received. Since bonuses can be substantial relative to monthly household income, our first stage estimates are sensitive to the inclusion of these payments.\(^\text{22}\)

Table 2 reports the estimates of the excess sensitivity of consumption with respect to income.\(^\text{23}\) Our full sample estimates are reported in the first column. Our OLS estimates of the impact of the marginal propensity of consume out of income, 0.086, while statistically significant, is somewhat small.

Since the $HH_{income}$ variable also contains many sources of income, some of which may have been unanticipated, we instead instrument for the change in monthly benefits received with the programmatic child benefit amount. We find a relatively large and significant estimate of 0.190 which typically would be interpreted as a marginal propensity to consume out of the income. This finding would usually be considered evidence of a substantial violation of the LCPIIH.

Turning to the first stage estimate, we see that the first stage estimate is 0.302. In the absence

\(^{22}\) Our first stage estimate when including bonus income is nearly twice the size of the corresponding estimate shown in Table 2. However, the result is still substantially less than one meaning that the IV estimates would still be biased upwards if we were to include bonus income in our measure of household income.

\(^{23}\) The standard errors are clustered at the household level.
of underreporting, we would expect this coefficient to equal one. That is, actual income should increase one for one with programmatic benefits. However, the large degree of underreporting attenuates the first stage estimate in the manner predicted by our analytical results. It is useful to note that if reported benefits were only affected by classical measurement error, we would anticipate that the estimated coefficient on the programmatic benefit would still equal one since left-hand side measurement error does not affect the slope parameters in the classical errors-in-variables model.

The reduced form estimate of 0.057 for the full sample shown in the final row of Table 2 is the consumption response to the programmatic change in benefits. In the absence of underreporting, we would expect the first stage estimate to equal one. Since the IV estimate is the ratio of the reduced form estimate to the first stage estimate, we would expect the IV estimate to simply equal the reduced form estimate if benefits were not underreported. Thus, the underreporting of child benefits inflates the causal estimate by roughly a factor of four. While the reduced form estimate still rejects the most basic LCPIH, the magnitude of the difference between this result and the above IV estimate yields a quite different substantive interpretation as to the deviation of behavior from the standard model.

One theoretical mechanism for why we might see excess sensitivity is the presence of liquidity constraints. Households that wish to borrow from future income but are unable to do so will respond to anticipated income changes since the marginal utility of a dollar today exceeds that of a dollar tomorrow. A common approach to testing for the importance of liquidity constraints is the split the sample between those that are likely to be constrained and those that are not (e.g., Zeldes 1989). Following in this tradition, we split the sample based upon whether the household is above or below the median income of child benefit eligible families in each survey year.

The 2SLS estimates shown in the final two columns of Table 2 show a large consumption response for both constrained (below median income) and unconstrained households. However, since the first stage estimate is only slightly larger for constrained households relative to unconstrained households, the reduced form findings are estimates of interest for the reasons discussed above. The reduced form estimate of 0.063 is significant for constrained group while the estimate of 0.049 is insignificant for the unconstrained group. However, these estimates are quite similar across the two groups. The key takeaway, however, is that the 2SLS estimates are dramatically overstated for
both groups due to underreporting of benefit receipt.

4 Conclusion

Survey non-response has continued to rise in recent years which has generated an increase in the share of observations with imputed values on a number of commonly used datasets. While the importance of this trend has been long recognized when using earnings, the share of imputed values for transfer benefits has received far less attention. In addition, benefits from a number of government programs are substantially underreported by respondents even after imputations are taken into account.

We demonstrate analytically that underreporting and imputation of government transfers can lead to a substantial overstatement of the impact of government transfers when applying instrumental variables methods to correct for the endogeneity and/or measurement error in these measures. Our theoretical predictions are borne out in both of our empirical examples. Our findings suggest that researchers investigating the efficacy of government transfers should follow the path routinely taken in the literatures examining wages gaps and returns to schooling and explicitly account for imputed benefit values in the estimation. Researchers can address this issue by assuming that benefits are missing at random in which case estimating the models on the non-imputed sample will yield consistent estimates, by explicitly modeling the missing data process, or even by developing bounds for the unknown parameters.

Our findings suggest that researchers should pay close attention to the magnitude of the first stage estimates in addition to the carefully examining the strength of the instruments. While the predicted magnitude of the first stage estimate is not always clear, as in the Social Security notch example, in many instances we have a strong prior regarding the first stage coefficients as in the child benefit example. Deviations from these priors should not simply be chalked up to measurement error but rather carefully investigated as the consequences for the consistency of the IV estimator could be quite severe.
Bibliography


<table>
<thead>
<tr>
<th>Sample:</th>
<th>Pooled S.S. Income</th>
<th>Non-Imputed S.S. Income</th>
<th>Imputed S.S. Income</th>
<th>Non-Imputed S.S. Income</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>(1)</td>
<td>(2)</td>
<td>(3)</td>
<td>(4)</td>
</tr>
<tr>
<td>A. Full Sample</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>OLS</td>
<td>-0.010</td>
<td>-0.012</td>
<td>-0.005</td>
<td>-0.012</td>
</tr>
<tr>
<td></td>
<td>(0.0004)</td>
<td>(0.0005)</td>
<td>(0.0007)</td>
<td>(0.0005)</td>
</tr>
<tr>
<td>First Stage</td>
<td>0.178</td>
<td>0.209</td>
<td>0.064</td>
<td>0.213</td>
</tr>
<tr>
<td>(Residuals)</td>
<td>(0.060)</td>
<td>(0.067)</td>
<td>(0.041)</td>
<td>(0.071)</td>
</tr>
<tr>
<td>Reduced Form</td>
<td>-0.0064</td>
<td>-0.0061</td>
<td>-0.0072</td>
<td>-0.0063</td>
</tr>
<tr>
<td>(Residuals)</td>
<td>(0.0029)</td>
<td>(0.0025)</td>
<td>(0.0047)</td>
<td>(0.0024)</td>
</tr>
<tr>
<td>2SLS</td>
<td>-0.036</td>
<td>-0.029</td>
<td>-0.137</td>
<td>-0.030</td>
</tr>
<tr>
<td></td>
<td>(0.022)</td>
<td>(0.017)</td>
<td>(0.117)</td>
<td>(0.016)</td>
</tr>
<tr>
<td>N</td>
<td>248,075</td>
<td>197,623</td>
<td>50,452</td>
<td>197,623</td>
</tr>
</tbody>
</table>

B. Widow Sample

| OLS     | -0.021             | -0.024                  | -0.012             | -0.023                  |
|         | (0.001)           | (0.001)               | (0.002)           | (0.001)               |
| First Stage | 0.104              | 0.124                   | 0.032              | 0.127                   |
| (Residuals) | (0.056)           | (0.064)                | (0.044)            | (0.064)                |
| Reduced Form | -0.015             | -0.014                  | -0.021             | -0.014                  |
| (Residuals) | (0.004)           | (0.004)               | (0.011)           | (0.003)               |
| 2SLS    | -0.146             | -0.111                  | -0.617             | -0.109                  |
|         | (0.087)           | (0.059)               | (0.811)           | (0.057)              |
| N       | 108,008            | 86,278                  | 21,730             | 86,278                  |

Notes: Each estimate in the Table is from a separate regression. The dependent variable is an indicator whether the family is living in a shared arrangement. The OLS and 2SLS estimates are the coefficients on family Social Security income and also include as controls: four categories each for the education of the head and the spouse (if present), age of the spouse (if present), marital status (married, widowed, and divorced), white, and female; a full set of indicators for the age (age+3 for widowed and divorced women) from ages 65 to 90; survey year indicators, and indicators for the nine Census divisions. The first stage and reduced form estimates are the coefficient on the Social Security instrument described in the text where the variables used in estimation are the residuals from regressions of the measures on the controls used in the OLS and 2SLS specifications. Standard errors are clustered at the year of birth.
<table>
<thead>
<tr>
<th>Sample:</th>
<th>Full Income</th>
<th>Below Median Income</th>
<th>Above Median Income</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>(1)</td>
<td>(2)</td>
<td>(3)</td>
</tr>
<tr>
<td>OLS</td>
<td>0.086</td>
<td>0.073</td>
<td>0.093</td>
</tr>
<tr>
<td></td>
<td>(0.010)</td>
<td>(0.007)</td>
<td>(0.015)</td>
</tr>
<tr>
<td>2SLS</td>
<td>0.190</td>
<td>0.189</td>
<td>0.176</td>
</tr>
<tr>
<td></td>
<td>(0.074)</td>
<td>(0.084)</td>
<td>(0.121)</td>
</tr>
<tr>
<td>First Stage</td>
<td>0.302</td>
<td>0.331</td>
<td>0.281</td>
</tr>
<tr>
<td></td>
<td>(0.032)</td>
<td>(0.044)</td>
<td>(0.046)</td>
</tr>
<tr>
<td>Reduced Form</td>
<td>0.057</td>
<td>0.063</td>
<td>0.049</td>
</tr>
<tr>
<td></td>
<td>(0.023)</td>
<td>(0.028)</td>
<td>(0.034)</td>
</tr>
<tr>
<td>N</td>
<td>217,312</td>
<td>108,391</td>
<td>108,921</td>
</tr>
</tbody>
</table>

Notes: Each estimate in the Table is from a separate regression. The dependent variable is the change in non-durable consumption from month \(t - 1\) to month \(t\). The OLS and 2SLS estimates are the coefficients on the change in reported other social security income from month \(t - 1\) to month \(t\). The first stage and reduced form estimates are the coefficient on the programmatic change in child benefits from month \(t - 1\) to month \(t\) which are computed based on the age and number of children in the household. Each regression also includes calendar year and month indicators, survey month indicators, the change in the number of household members, and the age of the household head and its square.
Figure 1: CPS Share Imputed Among Those With Positive Amounts 1988-2013

A. Earnings
B. Social Security
C. AFDC/TANF
D. Supplemental Security Inc.
E. Unemployment Insurance
F. Worker’s Compensation

Share Imputed

Year

1988
1998
2008

1988
1998
2008

1988
1998
2008

1988
1998
2008

1988
1998
2008

1988
1998
2008

1988
1998
2008

1988
1998
2008

1988
1998
2008

1988
1998
2008

1988
1998
2008
Figure 2: Social Security Income and the Instrument by Year of Birth

- S.S. Income: Non-Imputed

- S.S. Income: Imputed

- Instrument